Untitled

November 15, 2024

[1]:
```
import pandas as pd
import numpy as np

# Load datasets
movies = pd.read_csv("E:\Projects made by me\Movie Recommender␣
 ↪Systems\movies_metadata.csv", low_memory=False)
credits = pd.read_csv("E:\Projects made by me\Movie Recommender Systems\credits.
 ↪csv")
links = pd.read_csv("E:\Projects made by me\Movie Recommender Systems\links.
 ↪csv")
keywords = pd.read_csv("E:\Projects made by me\Movie Recommender␣
 ↪Systems\keywords.csv")
ratings = pd.read_csv(r"E:\Projects made by me\Movie Recommender␣
 ↪Systems\ratings_small.csv")

# Basic info about datasets to identify data types and missing values
print(movies.info())
print(credits.info())
print(links.info())
print(keywords.info())
print(ratings.info())
```

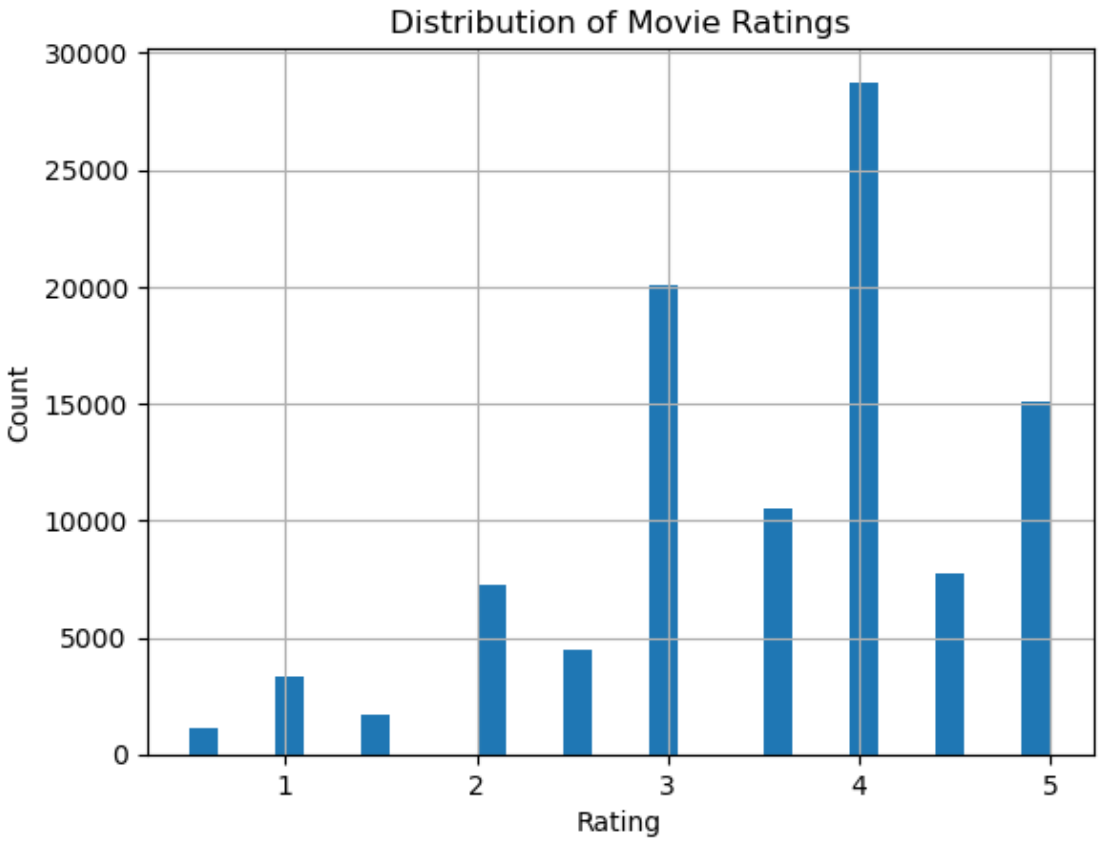
```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 45466 entries, 0 to 45465
Data columns (total 24 columns):
 #   Column                 Non-Null Count  Dtype
---  ------                 --------------  -----
 0   adult                  45466 non-null  object
 1   belongs_to_collection  4494 non-null   object
 2   budget                 45466 non-null  object
 3   genres                 45466 non-null  object
 4   homepage               7782 non-null   object
 5   id                     45466 non-null  object
 6   imdb_id                45449 non-null  object
 7   original_language      45455 non-null  object
 8   original_title         45466 non-null  object
 9   overview               44512 non-null  object
 10  popularity             45461 non-null  object
```

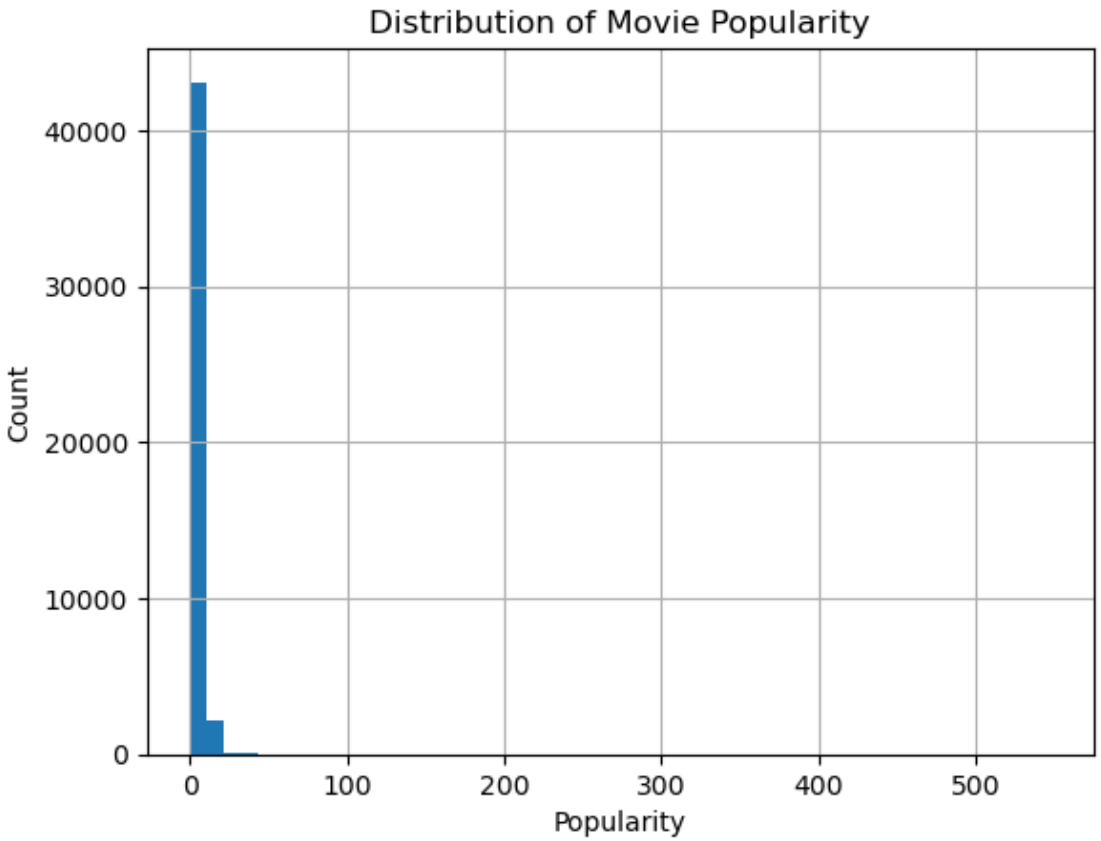
```
 11  poster_path            45080 non-null  object
 12  production_companies   45463 non-null  object
 13  production_countries   45463 non-null  object
 14  release_date           45379 non-null  object
 15  revenue                45460 non-null  float64
 16  runtime                45203 non-null  float64
 17  spoken_languages       45460 non-null  object
 18  status                 45379 non-null  object
 19  tagline                20412 non-null  object
 20  title                  45460 non-null  object
 21  video                  45460 non-null  object
 22  vote_average           45460 non-null  float64
 23  vote_count             45460 non-null  float64
dtypes: float64(4), object(20)
memory usage: 8.3+ MB
None
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 45476 entries, 0 to 45475
Data columns (total 3 columns):
 #   Column  Non-Null Count  Dtype
---  ------  --------------  -----
 0   cast    45476 non-null  object
 1   crew    45476 non-null  object
 2   id      45476 non-null  int64
dtypes: int64(1), object(2)
memory usage: 1.0+ MB
None
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 45843 entries, 0 to 45842
Data columns (total 3 columns):
 #   Column   Non-Null Count  Dtype
---  ------   --------------  -----
 0   movieId  45843 non-null  int64
 1   imdbId   45843 non-null  int64
 2   tmdbId   45624 non-null  float64
dtypes: float64(1), int64(2)
memory usage: 1.0 MB
None
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 46419 entries, 0 to 46418
Data columns (total 2 columns):
 #   Column    Non-Null Count  Dtype
---  ------    --------------  -----
 0   id        46419 non-null  int64
 1   keywords  46419 non-null  object
dtypes: int64(1), object(1)
memory usage: 725.4+ KB
None
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 100004 entries, 0 to 100003
Data columns (total 4 columns):
 #   Column     Non-Null Count   Dtype
---  ------     --------------   -----
 0   userId     100004 non-null  int64
 1   movieId    100004 non-null  int64
 2   rating     100004 non-null  float64
 3   timestamp  100004 non-null  int64
dtypes: float64(1), int64(3)
memory usage: 3.1 MB
None
```

[9]:
```python
# Convert 'popularity' column to numeric, handling errors by setting them to NaN
movies['popularity'] = pd.to_numeric(movies['popularity'], errors='coerce')

# Drop rows with essential missing values
movies.dropna(subset=['popularity'], inplace=True)

# Convert 'id' in movies to integer for easier merging with other datasets
movies['id'] = pd.to_numeric(movies['id'], errors='coerce')
movies.dropna(subset=['id'], inplace=True)
movies['id'] = movies['id'].astype(int)
```

[10]:
```python
import matplotlib.pyplot as plt
import seaborn as sns

# Plot the distribution of ratings
ratings['rating'].hist(bins=30)
plt.title('Distribution of Movie Ratings')
plt.xlabel('Rating')
plt.ylabel('Count')
plt.show()

# Plot movie popularity (example, if 'popularity' is numerical)
movies['popularity'].hist(bins=50)
plt.title('Distribution of Movie Popularity')
plt.xlabel('Popularity')
plt.ylabel('Count')
plt.show()
```


Distribution of Movie Ratings
Count
Rating
0
5000
10000
15000
20000
25000
30000
1
2
3
4
5



[11]:
```
# !pip install scikit-surprise
!pip show scikit-surprise
```

```
WARNING: Package(s) not found: scikit-surprise
```

[12]:
```
import sys
print(sys.executable)
```

```
C:\Users\DELL\anaconda3\python.exe
```

[13]:
```
!{sys.executable} -m pip install scikit-surprise
```

```
Collecting scikit-surprise
  Using cached scikit_surprise-1.1.4.tar.gz (154 kB)
  Installing build dependencies: started
  Installing build dependencies: finished with status 'done'
  Getting requirements to build wheel: started
  Getting requirements to build wheel: finished with status 'done'
  Preparing metadata (pyproject.toml): started
  Preparing metadata (pyproject.toml): finished with status 'done'
Requirement already satisfied: joblib>=1.2.0 in
c:\users\dell\anaconda3\lib\site-packages (from scikit-surprise) (1.2.0)
```

```
Requirement already satisfied: numpy>=1.19.5 in
c:\users\dell\anaconda3\lib\site-packages (from scikit-surprise) (1.24.3)
Requirement already satisfied: scipy>=1.6.0 in c:\users\dell\anaconda3\lib\site-
packages (from scikit-surprise) (1.11.1)
Building wheels for collected packages: scikit-surprise
  Building wheel for scikit-surprise (pyproject.toml): started
  Building wheel for scikit-surprise (pyproject.toml): finished with status
'error'
Failed to build scikit-surprise
  error: subprocess-exited-with-error

  Building wheel for scikit-surprise (pyproject.toml) did not run successfully.
  exit code: 1

  [115 lines of output]
  running bdist_wheel
  running build
  running build_py
  creating build\lib.win-amd64-cpython-311\surprise
  copying surprise\accuracy.py -> build\lib.win-amd64-cpython-311\surprise
  copying surprise\builtin_datasets.py -> build\lib.win-
amd64-cpython-311\surprise
  copying surprise\dataset.py -> build\lib.win-amd64-cpython-311\surprise
  copying surprise\dump.py -> build\lib.win-amd64-cpython-311\surprise
  copying surprise\reader.py -> build\lib.win-amd64-cpython-311\surprise
  copying surprise\trainset.py -> build\lib.win-amd64-cpython-311\surprise
  copying surprise\utils.py -> build\lib.win-amd64-cpython-311\surprise
  copying surprise\__init__.py -> build\lib.win-amd64-cpython-311\surprise
  copying surprise\__main__.py -> build\lib.win-amd64-cpython-311\surprise
  creating build\lib.win-amd64-cpython-311\surprise\model_selection
  copying surprise\model_selection\search.py -> build\lib.win-
amd64-cpython-311\surprise\model_selection
  copying surprise\model_selection\split.py -> build\lib.win-
amd64-cpython-311\surprise\model_selection
  copying surprise\model_selection\validation.py -> build\lib.win-
amd64-cpython-311\surprise\model_selection
  copying surprise\model_selection\__init__.py -> build\lib.win-
amd64-cpython-311\surprise\model_selection
  creating build\lib.win-amd64-cpython-311\surprise\prediction_algorithms
  copying surprise\prediction_algorithms\algo_base.py -> build\lib.win-
amd64-cpython-311\surprise\prediction_algorithms
  copying surprise\prediction_algorithms\baseline_only.py -> build\lib.win-
amd64-cpython-311\surprise\prediction_algorithms
  copying surprise\prediction_algorithms\knns.py -> build\lib.win-
amd64-cpython-311\surprise\prediction_algorithms
  copying surprise\prediction_algorithms\predictions.py -> build\lib.win-
amd64-cpython-311\surprise\prediction_algorithms
```

```
  copying surprise\prediction_algorithms\random_pred.py -> build\lib.win-
amd64-cpython-311\surprise\prediction_algorithms
  copying surprise\prediction_algorithms\__init__.py -> build\lib.win-
amd64-cpython-311\surprise\prediction_algorithms
  running egg_info
  writing scikit_surprise.egg-info\PKG-INFO
  writing dependency_links to scikit_surprise.egg-info\dependency_links.txt
  writing entry points to scikit_surprise.egg-info\entry_points.txt
  writing requirements to scikit_surprise.egg-info\requires.txt
  writing top-level names to scikit_surprise.egg-info\top_level.txt
  dependency C:\Users\DELL\AppData\Local\Temp\pip-build-env-
n_l_i7c9\overlay\Lib\site-packages\numpy\core\include\numpy\arrayobject.h won't
be automatically included in the manifest: the path must be relative
  dependency C:\Users\DELL\AppData\Local\Temp\pip-build-env-
n_l_i7c9\overlay\Lib\site-packages\numpy\core\include\numpy\arrayscalars.h won't
be automatically included in the manifest: the path must be relative
  dependency C:\Users\DELL\AppData\Local\Temp\pip-build-env-
n_l_i7c9\overlay\Lib\site-packages\numpy\core\include\numpy\ndarrayobject.h
won't be automatically included in the manifest: the path must be relative
  dependency C:\Users\DELL\AppData\Local\Temp\pip-build-env-
n_l_i7c9\overlay\Lib\site-packages\numpy\core\include\numpy\ndarraytypes.h won't
be automatically included in the manifest: the path must be relative
  dependency C:\Users\DELL\AppData\Local\Temp\pip-build-env-
n_l_i7c9\overlay\Lib\site-packages\numpy\core\include\numpy\ufuncobject.h won't
be automatically included in the manifest: the path must be relative
  dependency C:\Users\DELL\AppData\Local\Temp\pip-build-env-
n_l_i7c9\overlay\Lib\site-packages\numpy\core\include\numpy\arrayobject.h won't
be automatically included in the manifest: the path must be relative
  dependency C:\Users\DELL\AppData\Local\Temp\pip-build-env-
n_l_i7c9\overlay\Lib\site-packages\numpy\core\include\numpy\arrayscalars.h won't
be automatically included in the manifest: the path must be relative
  dependency C:\Users\DELL\AppData\Local\Temp\pip-build-env-
n_l_i7c9\overlay\Lib\site-packages\numpy\core\include\numpy\ndarrayobject.h
won't be automatically included in the manifest: the path must be relative
  dependency C:\Users\DELL\AppData\Local\Temp\pip-build-env-
n_l_i7c9\overlay\Lib\site-packages\numpy\core\include\numpy\ndarraytypes.h won't
be automatically included in the manifest: the path must be relative
  dependency C:\Users\DELL\AppData\Local\Temp\pip-build-env-
n_l_i7c9\overlay\Lib\site-packages\numpy\core\include\numpy\ufuncobject.h won't
be automatically included in the manifest: the path must be relative
  dependency C:\Users\DELL\AppData\Local\Temp\pip-build-env-
n_l_i7c9\overlay\Lib\site-packages\numpy\core\include\numpy\arrayobject.h won't
be automatically included in the manifest: the path must be relative
  dependency C:\Users\DELL\AppData\Local\Temp\pip-build-env-
n_l_i7c9\overlay\Lib\site-packages\numpy\core\include\numpy\arrayscalars.h won't
be automatically included in the manifest: the path must be relative
  dependency C:\Users\DELL\AppData\Local\Temp\pip-build-env-
n_l_i7c9\overlay\Lib\site-packages\numpy\core\include\numpy\ndarrayobject.h
```

```
won't be automatically included in the manifest: the path must be relative
  dependency C:\Users\DELL\AppData\Local\Temp\pip-build-env-
n_l_i7c9\overlay\Lib\site-packages\numpy\core\include\numpy\ndarraytypes.h won't
be automatically included in the manifest: the path must be relative
  dependency C:\Users\DELL\AppData\Local\Temp\pip-build-env-
n_l_i7c9\overlay\Lib\site-packages\numpy\core\include\numpy\ufuncobject.h won't
be automatically included in the manifest: the path must be relative
  dependency C:\Users\DELL\AppData\Local\Temp\pip-build-env-
n_l_i7c9\overlay\Lib\site-packages\numpy\core\include\numpy\arrayobject.h won't
be automatically included in the manifest: the path must be relative
  dependency C:\Users\DELL\AppData\Local\Temp\pip-build-env-
n_l_i7c9\overlay\Lib\site-packages\numpy\core\include\numpy\arrayscalars.h won't
be automatically included in the manifest: the path must be relative
  dependency C:\Users\DELL\AppData\Local\Temp\pip-build-env-
n_l_i7c9\overlay\Lib\site-packages\numpy\core\include\numpy\ndarrayobject.h
won't be automatically included in the manifest: the path must be relative
  dependency C:\Users\DELL\AppData\Local\Temp\pip-build-env-
n_l_i7c9\overlay\Lib\site-packages\numpy\core\include\numpy\ndarraytypes.h won't
be automatically included in the manifest: the path must be relative
  dependency C:\Users\DELL\AppData\Local\Temp\pip-build-env-
n_l_i7c9\overlay\Lib\site-packages\numpy\core\include\numpy\ufuncobject.h won't
be automatically included in the manifest: the path must be relative
  dependency C:\Users\DELL\AppData\Local\Temp\pip-build-env-
n_l_i7c9\overlay\Lib\site-packages\numpy\core\include\numpy\arrayobject.h won't
be automatically included in the manifest: the path must be relative
  dependency C:\Users\DELL\AppData\Local\Temp\pip-build-env-
n_l_i7c9\overlay\Lib\site-packages\numpy\core\include\numpy\arrayscalars.h won't
be automatically included in the manifest: the path must be relative
  dependency C:\Users\DELL\AppData\Local\Temp\pip-build-env-
n_l_i7c9\overlay\Lib\site-packages\numpy\core\include\numpy\ndarrayobject.h
won't be automatically included in the manifest: the path must be relative
  dependency C:\Users\DELL\AppData\Local\Temp\pip-build-env-
n_l_i7c9\overlay\Lib\site-packages\numpy\core\include\numpy\ndarraytypes.h won't
be automatically included in the manifest: the path must be relative
  dependency C:\Users\DELL\AppData\Local\Temp\pip-build-env-
n_l_i7c9\overlay\Lib\site-packages\numpy\core\include\numpy\ufuncobject.h won't
be automatically included in the manifest: the path must be relative
  reading manifest file 'scikit_surprise.egg-info\SOURCES.txt'
  reading manifest template 'MANIFEST.in'
  warning: no previously-included files matching '*.so' found under directory
'surprise'
  adding license file 'LICENSE.md'
  writing manifest file 'scikit_surprise.egg-info\SOURCES.txt'
  C:\Users\DELL\AppData\Local\Temp\pip-build-env-n_l_i7c9\overlay\Lib\site-
packages\setuptools\command\build_py.py:219: _Warning: Package
'surprise.prediction_algorithms' is absent from the `packages` configuration.
  !!
```

```
**********************************************************************************
          ############################
          # Package would be ignored #
          ############################
          Python recognizes 'surprise.prediction_algorithms' as an importable
package[^1],
          but it is absent from setuptools' `packages` configuration.

          This leads to an ambiguous overall configuration. If you want to
distribute this
          package, please make sure that 'surprise.prediction_algorithms' is
explicitly added
          to the `packages` configuration field.

          Alternatively, you can also rely on setuptools' discovery methods
          (for example by using `find_namespace_packages(…)`/`find_namespace:`
          instead of `find_packages(…)`/`find:`).

          You can read more about "package discovery" on setuptools
documentation page:

          -
https://setuptools.pypa.io/en/latest/userguide/package_discovery.html

          If you don't want 'surprise.prediction_algorithms' to be distributed
and are
          already explicitly excluding 'surprise.prediction_algorithms' via
          `find_namespace_packages(…)/find_namespace` or
`find_packages(…)/find`,
          you can try to use `exclude_package_data`, or `include-package-
data=False` in
          combination with a more fine grained `package-data` configuration.

          You can read more about "package data files" on setuptools
documentation page:

          - https://setuptools.pypa.io/en/latest/userguide/datafiles.html


          [^1]: For Python, any directory (with suitable naming) can be
imported,
                even if it does not contain any `.py` files.
                On the other hand, currently there is no concept of package data
                directory, all directories are treated like packages.
**********************************************************************************

  !!
    check.warn(importable)
```

```
  copying surprise\similarities.c -> build\lib.win-amd64-cpython-311\surprise
  copying surprise\similarities.pyx -> build\lib.win-amd64-cpython-311\surprise
  copying surprise\prediction_algorithms\co_clustering.c -> build\lib.win-
amd64-cpython-311\surprise\prediction_algorithms
  copying surprise\prediction_algorithms\matrix_factorization.c ->
build\lib.win-amd64-cpython-311\surprise\prediction_algorithms
  copying surprise\prediction_algorithms\optimize_baselines.c -> build\lib.win-
amd64-cpython-311\surprise\prediction_algorithms
  copying surprise\prediction_algorithms\slope_one.c -> build\lib.win-
amd64-cpython-311\surprise\prediction_algorithms
  copying surprise\prediction_algorithms\co_clustering.pyx -> build\lib.win-
amd64-cpython-311\surprise\prediction_algorithms
  copying surprise\prediction_algorithms\matrix_factorization.pyx ->
build\lib.win-amd64-cpython-311\surprise\prediction_algorithms
  copying surprise\prediction_algorithms\optimize_baselines.pyx ->
build\lib.win-amd64-cpython-311\surprise\prediction_algorithms
  copying surprise\prediction_algorithms\slope_one.pyx -> build\lib.win-
amd64-cpython-311\surprise\prediction_algorithms
  running build_ext
  building 'surprise.similarities' extension
  error: Microsoft Visual C++ 14.0 or greater is required. Get it with
"Microsoft C++ Build Tools": https://visualstudio.microsoft.com/visual-cpp-
build-tools/
  [end of output]

  note: This error originates from a subprocess, and is likely not a problem
with pip.
  ERROR: Failed building wheel for scikit-surprise
ERROR: Could not build wheels for scikit-surprise, which is required to install
pyproject.toml-based projects
```

[14]:
```python
from surprise import Dataset, Reader, SVD
from surprise.model_selection import cross_validate

# Convert ratings into Surprise format
reader = Reader(rating_scale=(0.5, 5))
data = Dataset.load_from_df(ratings[['userId', 'movieId', 'rating']], reader)

# Define the SVD model
svd = SVD()

# Perform cross-validation
cross_validate(svd, data, measures=['RMSE', 'MAE'], cv=5, verbose=True)
```

```
---------------------------------------------------------------------------
ModuleNotFoundError                       Traceback (most recent call last)
Cell In[14], line 1
```

```
----> 1 from surprise import Dataset, Reader, SVD
      2 from surprise.model_selection import cross_validate
      4 # Convert ratings into Surprise format

ModuleNotFoundError: No module named 'surprise'
```

[ ]:

[19]:
```
trainset = data.build_full_trainset()
svd.fit(trainset)
```

```
---------------------------------------------------------------------------
NameError                                 Traceback (most recent call last)
Cell In[19], line 1
----> 1 trainset = data.build_full_trainset()
      2 svd.fit(trainset)

NameError: name 'data' is not defined
```

[20]:
```
# Predict the rating for a user (ID=1) for a specific movie (ID=10)
user_id = 1
movie_id = 10
prediction = svd.predict(user_id, movie_id)
print(f"Predicted rating for user {user_id} for movie {movie_id}: {prediction.
 ↪est}")
```

```
---------------------------------------------------------------------------
NameError                                 Traceback (most recent call last)
Cell In[20], line 4
      2 user_id = 1
      3 movie_id = 10
----> 4 prediction = svd.predict(user_id, movie_id)
      5 print(f"Predicted rating for user {user_id} for movie {movie_id}:␣
 ↪{prediction.est}")

NameError: name 'svd' is not defined
```

[21]:
```
def recommend_movies(user_id, num_recommendations=10):
    # Extract unique movie IDs from the dataset
    movie_ids = movies['id'].unique()

    # Predict ratings for each movie that the user hasn't rated
    movie_ratings = [svd.predict(user_id, movie_id).est for movie_id in␣
 ↪movie_ids]
```

```
    # Create a DataFrame with movie IDs and predicted ratings
    recommendations = pd.DataFrame({
        'movieId': movie_ids,
        'predicted_rating': movie_ratings
    }).sort_values(by='predicted_rating', ascending=False)

    # Get the top N recommended movies
    top_recommendations = recommendations.head(num_recommendations)
    top_recommendations = top_recommendations.merge(movies[['id', 'title']],␣
 ↪left_on='movieId', right_on='id')

    return top_recommendations[['title', 'predicted_rating']]

# Example: Recommend top 10 movies for user ID=1
recommendations = recommend_movies(1, 10)
print(recommendations)
```

```
---------------------------------------------------------------------------
NameError                                 Traceback (most recent call last)
Cell In[21], line 21
     18     return top_recommendations[['title', 'predicted_rating']]
     20 # Example: Recommend top 10 movies for user ID=1
---> 21 recommendations = recommend_movies(1, 10)
     22 print(recommendations)

Cell In[21], line 6, in recommend_movies(user_id, num_recommendations)
      3 movie_ids = movies['id'].unique()
      5 # Predict ratings for each movie that the user hasn't rated
----> 6 movie_ratings = [svd.predict(user_id, movie_id).est for movie_id in␣
 ↪movie_ids]
      8 # Create a DataFrame with movie IDs and predicted ratings
      9 recommendations = pd.DataFrame({
     10     'movieId': movie_ids,
     11     'predicted_rating': movie_ratings
     12 }).sort_values(by='predicted_rating', ascending=False)

Cell In[21], line 6, in <listcomp>(.0)
      3 movie_ids = movies['id'].unique()
      5 # Predict ratings for each movie that the user hasn't rated
----> 6 movie_ratings = [svd.predict(user_id, movie_id).est for movie_id in␣
 ↪movie_ids]
      8 # Create a DataFrame with movie IDs and predicted ratings
      9 recommendations = pd.DataFrame({
     10     'movieId': movie_ids,
     11     'predicted_rating': movie_ratings
     12 }).sort_values(by='predicted_rating', ascending=False)
```

```
NameError: name 'svd' is not defined
```

[22]:
```
from flask import Flask, request, jsonify

app = Flask(__name__)

@app.route('/recommend', methods=['POST'])
def recommend():
    data = request.get_json(force=True)
    user_id = data['user_id']
    num_recommendations = data.get('num_recommendations', 10)
    recommendations = recommend_movies(user_id, num_recommendations)
    return jsonify(recommendations.to_dict(orient='records'))

if __name__ == '__main__':
    app.run(debug=True)
```

```
 * Serving Flask app '__main__'
 * Debug mode: on
WARNING: This is a development server. Do not use it in a production deployment.
Use a production WSGI server instead.
 * Running on http://127.0.0.1:5000
Press CTRL+C to quit
 * Restarting with watchdog (windowsapi)
```

```
An exception has occurred, use %tb to see the full traceback.

SystemExit: 1
```

```
C:\Users\DELL\anaconda3\Lib\site-packages\IPython\core\interactiveshell.py:3534:
UserWarning: To exit: use 'exit', 'quit', or Ctrl-D.
  warn("To exit: use 'exit', 'quit', or Ctrl-D.", stacklevel=1)
```

[ ]: